TOWARDS A SAMPLING THEOREM FOR SIGNALS ON ARBITRARY GRAPHS

Aamir Anis, Akshay Gadde and Antonio Ortega

Department of Electrical Engineering
University of Southern California, Los Angeles
Email: aanis@usc.edu, agadde@usc.edu, ortega@sipi.usc.edu

ABSTRACT

In this paper, we extend the Nyquist-Shannon theory of sampling to signals defined on arbitrary graphs. Using spectral graph theory, we establish a cut-off frequency for all bandlimited graph signals that can be perfectly reconstructed from samples on a given subset of nodes. The result is analogous to the concept of Nyquist frequency in traditional signal processing. We consider practical ways of computing this cut-off and show that it is an improvement over previous results. We also propose a greedy algorithm to search for the smallest possible sampling set that guarantees unique recovery for a signal of given bandwidth. The efficacy of these results is verified through simple examples.

Index Terms— Graph signal processing, sampling theorem, spectral graph theory

1. INTRODUCTION

Representation, processing and analysis of large-scale data as signals defined on graphs has drawn much interest recently. Graphs allow us to embed natural inter-connectivities between data points and exploit them during processing. As a result, graph signal processing has laid a strong foothold in various application domains such as machine learning [1], sensor networks [2] and image processing [3, 4, 5]. Although powerful, this research area is still in its infancy. Recent efforts have therefore focused on translating well-developed tools of traditional signal processing for handling graph signals (see [6] for a comprehensive overview).

Sampling theory is one aspect of graph signal processing that is still not fully understood. In traditional signal processing, given a signal of bandwidth f and a (uniform) sampling rate f_s, the Nyquist-Shannon sampling theorem gives the condition for unique recovery of the signal from its samples as $f < f_s/2$. If the bandwidth is given, the required sampling rate can be easily calculated and vice-versa. Analogous results relating the bandwidth and the required sampling density have also been proposed for irregular sampling [7]. In graph signals, the notion of frequency is introduced via the eigenvalues and eigenvectors of the graph Laplacian. Posing a sampling theorem in this context is challenging because graph signals do not necessarily lie in regular Euclidean spaces. For example, downsampling in discrete-time signals involves dropping every other sample. Moreover, such downsampling leads to a frequency folding phenomenon that we can leverage to state the sampling theorem. However, we cannot define such a sampling pattern for arbitrary graphs, i.e., dropping "every other vertex" is not a well-defined notion in this context. Naturally, to frame a *sampling theorem* for graphs, we need to consider the following questions. Firstly, what is the maximum possible bandwidth (the cut-off frequency) of a graph signal such that it is uniquely sampled onto a given subset of nodes, and conversely, what is the smallest possible subset of nodes on which a signal of given bandwidth is uniquely represented by its samples.

Sampling theory for graph signals has been studied before. In the case of bipartite graphs, downsampling on one of the colored partitions leads to an effect analogous to frequency folding [8]. This gives the cut-off frequency and also suggests a natural sampling strategy. For arbitrary graphs, [9] gives a sufficient condition that the sampling set needs to satisfy for unique recovery. Using this condition, a bound on the maximum bandwidth of all recoverable signals is given in [10]. The problem of downsampling on arbitrary graphs is considered in [11] in the context of multi-scale transforms. However, this work does not give a method to search for sampling sets that guarantee unique recovery of signals of given bandwidth.

In this work, we address both the questions required for a complete sampling theorem. We first present a *necessary and sufficient* condition for unique recovery of a bandlimited signal from its samples. We then provide a novel, computationally tractable method of computing the cut-off frequency for a given sampling set and show that our result is an improvement over previous work. Additionally, we provide a novel algorithm for choosing the smallest possible sampling set for a given bandwidth using a greedy heuristic. The rest of the paper is organized as follows: in Section 2, after briefly reviewing prior work, we formulate the sampling theorem for graph signals and give practical ways of using it. In Section 3, we provide simple experiments to back our proposed theory and conclude with future directions in Section 4.

2. SAMPLING THEORY IN GRAPHS

2.1. Background and notation

We start by introducing the notations used throughout this paper. A simple, connected, undirected, and weighted graph $G = (\mathcal{V}, E)$ consists of a set of nodes $\mathcal{V} = \{1, 2, \dots, N\}$ and edges $E = \{(i, j, w_{ij})\}, i, j \in \mathcal{V}$, where (i, j, w_{ij}) denotes an edge of weight w_{ij} between node i and j, with $w_{ii} = 0$. The degree d_i of a node i is the sum of the edge-weights connected to node i, and the degree matrix of the graph consists of degrees of all the nodes arranged in a diagonal matrix $\mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_N\}$. The adjacency matrix $\mathbf{W}$ of the graph is an $N \times N$ matrix with $\mathbf{W}(i, j) = w_{ij}$ and the Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{W}$. We shall use the symmetric normalized form of the adjacency and the Laplacian matrices defined as $\boldsymbol{\mathcal{W}} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ and $\boldsymbol{\mathcal{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$ respectively. $\boldsymbol{\mathcal{L}}$ is a symmetric positive semi-definite matrix and has a set of real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \leq 2$ and a corresponding orthogonal set of eigenvectors denoted as $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N\}$. We denote a subset of nodes of the graph as a collection of indices $\mathcal{S} \subset \mathcal{V}$, with $\mathcal{S}^c = \mathcal{V} \setminus \mathcal{S}$ denoting its complement set. A restriction of a matrix $\mathbf{A}$ to rows in set $\mathcal{S}_1$ and columns in set $\mathcal{S}_2$

This work was supported in part by NSF under grant CCF-1018977.

is denoted by the submatrix $(\mathbf{A})_{\mathcal{S}_1,\mathcal{S}_2}$ and for the sake of brevity $\mathbf{A}_{\mathcal{S},\mathcal{S}} = \mathbf{A}_{\mathcal{S}}$. A graph signal is defined as a scalar valued discrete mapping $f : \mathcal{V} \to \mathbb{R}$, such that $f(i)$ is the value of the signal on node i. Thus a graph signal can also be represented as a vector $\mathbf{f}$ in $\mathbb{R}^N$, with indices corresponding to the nodes in the graph. The *downsampling* operation on a graph signal $\mathbf{f}$ is defined as the restriction of the signal $\mathbf{f}$ to a certain subset of nodes $\mathcal{S}$ known as the *downsampling set*, and the downsampled signal is denoted by $\mathbf{f}(\mathcal{S})$, which is a vector of reduced length $|\mathcal{S}|$.

It is known that the eigenvalues and eigenvectors of $\boldsymbol{\mathcal{L}}$ provide a spectral interpretation (i.e. a notion of frequency) for a graph signal, similar to the Fourier transform in traditional signal processing. The eigenvalues of $\boldsymbol{\mathcal{L}}$ can be thought of as frequencies - a high eigenvalue results in higher variation in the corresponding eigenvector [6]. The *graph Fourier transform* (GFT) of a signal $\mathbf{f}$ is defined as its projection onto the eigenvectors of the graph Laplacian, i.e. $\tilde{\mathbf{f}}(\lambda_i) = \langle \mathbf{f}, \mathbf{u}_i \rangle$, or more compactly, $\tilde{\mathbf{f}} = \mathbf{U}^T\mathbf{f}$. In this setting, an *$\omega$-bandlimited signal* on a graph is defined to have zero GFT coefficients for frequencies above its bandwidth ω, i.e. its spectral support is restricted to the set of frequencies $[0, \omega]$. The space of all ω-bandlimited signals is known as the *Paley-Wiener* space and is denoted by $PW_\omega(G)$ [9]. Note that $PW_\omega(G)$ is a subspace of $\mathbb{R}^N$.

2.2. Prior work

To formulate the sampling problem for arbitrary graphs, it is useful to state the concept of *uniqueness sets* [9]:

Definition 1 (Uniqueness set)**.** *A subset of nodes $\mathcal{S} \subset \mathcal{V}$ is a uniqueness set for the space $PW_\omega(G)$, if for any two signals from $PW_\omega(G)$, the fact that they coincide on $\mathcal{S}$ implies that they coincide on $\mathcal{V}$, i.e., $\forall\, \mathbf{f}, \mathbf{g} \in PW_\omega(G), \mathbf{f}(\mathcal{S}) = \mathbf{g}(\mathcal{S}) \Rightarrow \mathbf{f} = \mathbf{g}$*

The definition above implies that sampling onto the set $\mathcal{S}$ is a one-to-one operation for an ω-bandlimited signal and thus it is sufficient to know its values on the uniqueness set $\mathcal{S}$. Based on this definition and the work of [9], the following theorem was proposed in [10] to estimate the cut-off:

Theorem 1. *For a graph G with normalized Laplacian matrix $\boldsymbol{\mathcal{L}}$, a set $\mathcal{S}$ is a uniqueness set for all signals $\mathbf{f} \in PW_\omega(G)$ if $\omega < \omega^*_{\mathcal{S}}$ with $\omega^*_{\mathcal{S}} = \sigma_{\min}$, where $\sigma^2_{\min}$ is the smallest eigenvalue of $(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}$.*

This theorem gives only a *sufficient* condition and does not guarantee that the estimated cutoff-frequency is the best possible, given $\mathcal{S}$. Therefore, we would like to give a necessary and sufficient condition for $\mathcal{S}$ to be a uniqueness set and provide a practical way to improve the estimate of the cut-off frequency.

The work of [11] suggests a method for graph downsampling, based on the polarity of components of the largest Laplacian eigenvector, for designing two-channel filterbanks. However, this method does not address the downsampling problem in the context of the sampling theorem, namely, choosing sampling sets that guarantee unique recovery of signals of *given* bandwidth.

2.3. Cut-off frequency

To derive a stronger condition for a subset of nodes to be a uniqueness set for ω-bandlimited signals, we observe the following:

Lemma 1. *$\mathcal{S}$ is a uniqueness set for signals in $PW_\omega(G)$ if and only if $PW_\omega(G) \cap L_2(\mathcal{S}^c) = \{\mathbf{0}\}$, where $L_2(\mathcal{S}^c)$ is the space of all graph signals that are zero everywhere except possibly on the subset of nodes $\mathcal{S}^c$, i.e., $\forall \phi \in L_2(\mathcal{S}^c), \phi(\mathcal{S}) = 0$.*

Proof. Given $PW_\omega(G) \cap L_2(\mathcal{S}^c) = \{\mathbf{0}\}$, assume that $\mathcal{S}$ is not a uniqueness set. Then, there exist $\mathbf{f}, \mathbf{g} \in PW_\omega(G), \mathbf{f} \neq \mathbf{g}$ such that $\mathbf{f}(\mathcal{S}) = \mathbf{g}(\mathcal{S})$. Hence, we have $\mathbf{f} - \mathbf{g} \in L_2(\mathcal{S}^c), \mathbf{f} - \mathbf{g} \neq \mathbf{0}$. Also, $\mathbf{f} - \mathbf{g} \in PW_\omega(G)$ due to closure. But this is a contradiction as $PW_\omega(G) \cap L_2(\mathcal{S}^c) = \{\mathbf{0}\}$. Therefore $\mathcal{S}$ must be a uniqueness set.

Conversely, we are given that $\mathcal{S}$ is a uniqueness set. Let ϕ be any signal in $PW_\omega(G) \cap L_2(\mathcal{S}^c)$. Then, for any $\mathbf{f} \in PW_\omega(G)$, we have $\mathbf{g} = \mathbf{f} + \phi \in PW_\omega(G)$ and $\mathbf{f}(\mathcal{S}) = \mathbf{g}(\mathcal{S})$. But since $\mathcal{S}$ is a uniqueness set, one must have $\mathbf{f} = \mathbf{g}$, which implies $\phi = \mathbf{0}$. Therefore, $PW_\omega(G) \cap L_2(\mathcal{S}^c) = \{\mathbf{0}\}$.

The lemma above gives us various insights. Firstly, we note that the set $L_2(\mathcal{S}^c)$ is equal to the null space $\mathcal{N}(\mathcal{D}_{\mathcal{S}})$ of the downsampling operator $\mathcal{D}_{\mathcal{S}} : \mathbb{R}^N \to \mathbb{R}^{|\mathcal{S}|}$, since $\phi(\mathcal{S}) = \mathbf{0}, \forall \phi \in L_2(\mathcal{S}^c)$. Enforcing $PW_\omega(G) \cap \mathcal{N}(\mathcal{D}_{\mathcal{S}}) = \{\mathbf{0}\}$ ensures that downsampling signals from $PW_\omega(G)$ onto the set $\mathcal{S}$ leads to a one-to-one mapping, and thus allows unique reconstruction. Secondly, we get a hint on estimating the cut-off frequency for signals such that they have $\mathcal{S}$ as a uniqueness set – choose $PW_\omega(G)$ such that ω is less than the minimum possible bandwidth of all signals in $L_2(\mathcal{S}^c)$. This would ensure that no signal from $L_2(\mathcal{S}^c)$ can be a part of $PW_\omega(G)$. Let $\omega(\mathbf{f})$ be the bandwidth of a signal $\mathbf{f}$ (i.e. the largest among eigenvalues corresponding to non-zero GFT coefficients of $\mathbf{f}$). We now state the following theorem:

Theorem 2 (Sampling Theorem)**.** *For a graph G, with normalized Laplacian $\boldsymbol{\mathcal{L}}$, a subset of nodes $\mathcal{S} \subset \mathcal{V}$ is a uniqueness set for signals $\mathbf{f} \in PW_\omega(G)$ if and only if*

$$\omega < \inf_{\phi \in L_2(\mathcal{S}^c)} \omega(\phi) \triangleq \omega_c(\mathcal{S}) \tag{1}$$

We call $\omega_c(\mathcal{S})$ the true cut-off frequency.

To use the theorem above, we first need a tool to compute the bandwidth $\omega(\phi)$ of any given signal ϕ. This can be done trivially using the GFT. However, since we also need to minimize the bandwidth over all signals in $L_2(\mathcal{S}^c)$, we propose an alternate method for bandwidth estimation by defining the following quantity:

Definition 2 (Spectral moments)**.** *For any signal $\mathbf{f} \neq \mathbf{0}$, we define its k^{th} spectral moment $\omega_k(\mathbf{f})$ with $k \in \mathbb{Z}^+$ as*

$$\omega_k(\mathbf{f}) \triangleq \left(\frac{\mathbf{f}^t \boldsymbol{\mathcal{L}}^k \mathbf{f}}{\mathbf{f}^t \mathbf{f}}\right)^{1/k} = \left(\sum_{i=1}^{N} \lambda_i^k \frac{\tilde{\mathbf{f}}(i)^2}{\sum_{j=1}^{N} \tilde{\mathbf{f}}(j)^2}\right)^{1/k} \tag{2}$$

We call these quantities "spectral moments" because they are the moments of graph frequencies with respect to the signal's energy distribution. $\omega_k(\mathbf{f})$ can be shown to increase monotonically with k:

$$\forall \mathbf{f}, k_1 < k_2 \Rightarrow \omega_{k_1}(\mathbf{f}) \leq \omega_{k_2}(\mathbf{f}) \tag{3}$$

These moments are bounded from above, hence $\lim_{k\to\infty} \omega_k(\mathbf{f})$ exists for all $\mathbf{f}$. Consequently, it is easy to prove that if $\omega(\mathbf{f})$ denotes the bandwidth of a signal $\mathbf{f}$, then

$$\forall k > 0,\ \omega_k(\mathbf{f}) \leq \lim_{j\to\infty} \omega_j(\mathbf{f}) = \omega(\mathbf{f}) \tag{4}$$

This gives us an important insight: the spectral moment of a signal, for a finite but large k, has a value close to (but less than) the actual bandwidth of the signal, i.e., it essentially indicates the frequency localization of the signal energy. Therefore, using $\omega_k(\phi)$ as a proxy

for $\omega(\phi)$ (i.e. bandwidth of ϕ) is justified and this leads us to define the *cut-off frequency estimate of order k* as

$$\Omega_k(\mathcal{S}) \triangleq \inf_{\phi \in L_2(\mathcal{S}^c)} \omega_k(\phi) = \inf_{\phi \in L_2(\mathcal{S}^c)} \left(\frac{\phi^t \boldsymbol{\mathcal{L}}^k \phi}{\phi^t \phi} \right)^{1/k} \tag{5}$$

We note that this cut-off has a form analogous to the *minimum Dirichlet eigenvalue* of the sub-graph $\mathcal{S}^c$ [12]. Using the definitions of $\Omega_k(\mathcal{S})$ and $\omega_c(\mathcal{S})$ along with (3) and (4), we conclude that for any $k_1 < k_2$:

$$\omega_c(\mathcal{S}) \geq \lim_{k \to \infty} \Omega_k(\mathcal{S}) \geq \Omega_{k_2}(\mathcal{S}) \geq \Omega_{k_1}(\mathcal{S}) \tag{6}$$

This equation has significant implications: Firstly, as we increase the value of k, $\Omega_k(\mathcal{S})$ tends to give a better estimate of the cut-off frequency. Note that $\Omega_2(\mathcal{S})$ is exactly equal to the cut-off derived in Theorem 1 - this means for any $k > 2$, the estimated cut-off is better than previously stated in [10]. Secondly, $\Omega_k(\mathcal{S})$ is always less than the actual cutoff $\omega_c(\mathcal{S})$, thus ensuring stable reconstruction. This is because if $\mathcal{S}$ is a uniqueness set for $PW_{\omega_1}(G)$ and $\omega_1 > \omega_2$, then $PW_{\omega_2}(G) \subseteq PW_{\omega_1}(G)$ and hence $\mathcal{S}$ is also a uniqueness set for $PW_{\omega_2}(G)$ (from Lemma 1). Using (6) and Theorem 2, we give the following proposition:

Proposition 1. *For all k, $\mathcal{S} \subset \mathcal{V}$ is a uniqueness set for $PW_\omega(G)$ if $\omega < \Omega_k(\mathcal{S})$. $\Omega_k(\mathcal{S})$ can be computed from (5) as*

$$\Omega_k(\mathcal{S}) = \left[\inf_{\psi} \frac{\psi^t (\boldsymbol{\mathcal{L}}^k)_{\mathcal{S}^c} \psi}{\psi^t \psi} \right]^{1/k} = (\sigma_{1,k})^{1/k} \tag{7}$$

where $\sigma_{1,k}$ denotes the smallest eigenvalue of the reduced matrix $(\boldsymbol{\mathcal{L}}^k)_{\mathcal{S}^c}$. Further, if $\psi_{1,k}$ is the corresponding eigenvector, and ϕ_k^* minimizes $\omega_k(\phi)$ in (5) (i.e. it approximates the smoothest possible signal in $L_2(\mathcal{S}^c)$), then

$$\phi_k^*(\mathcal{S}^c) = \psi_{1,k}, \quad \phi_k^*(\mathcal{S}) = \mathbf{0} \tag{8}$$

Finally, we also note from (6) that one simply needs a higher k to get a better estimate of the true cut-off frequency. Therefore, there is a trade-off between accuracy of the estimate on the one hand, and complexity and numerical stability on the other (that arise due to the term $\boldsymbol{\mathcal{L}}^k$). Once we have an estimate of the cut-off, interpolation can be performed using techniques proposed in [10, 13, 14].

2.4. Smallest sampling set

Now, we turn to the converse question: given a signal $\mathbf{f}$ of certain bandwidth ω_c, what is the smallest set $\mathcal{S}_{\text{opt}}$ so that the signal is uniquely represented by $\mathbf{f}(\mathcal{S}_{\text{opt}})$. If K_c represents the number of eigenvalues of $\boldsymbol{\mathcal{L}}$ below ω_c, then by dimensionality considerations $|\mathcal{S}_{\text{opt}}| \geq K_c$. Also, note that $\mathcal{S}_{\text{opt}}$ may not be unique. Formally, if we use Theorem 2 and relax the true cut-off $\omega_c(\mathcal{S})$ by $\Omega_k(\mathcal{S})$, then $\mathcal{S}_{\text{opt}}$ can be found from the following optimization problem:

$$\underset{\mathcal{S}}{\text{Minimize}} \ \ |\mathcal{S}| \quad \text{subject to} \quad \Omega_k(\mathcal{S}) \geq \omega_c \tag{9}$$

This is a combinatorial problem because we need to compute $\Omega_k(\mathcal{S})$ for every possible subset $\mathcal{S}$. We therefore formulate a greedy heuristic to get an estimate $\mathcal{S}_{\text{est}}$ of the optimal sampling set. Starting with an empty sampling set $\mathcal{S}$ ($\Omega_k(\mathcal{S}) = 0$) we keep adding nodes (from $\mathcal{S}^c$) one-by-one while trying to ensure maximum increase in $\Omega_k(\mathcal{S})$ at each step. The hope is that $\Omega_k(\mathcal{S})$ reaches the target cut-off ω_c with minimum number of node additions to $\mathcal{S}$. To achieve this, we first tackle the combinatorial nature of our problem by defining the following matrix

Algorithm 1 Greedy heuristic for estimating $\mathcal{S}_{\text{opt}}$

Input: $G = \{\mathcal{V}, E\}$, $\boldsymbol{\mathcal{L}}$, bandwidth ω_c, some $k \in \mathbb{Z}^+$
Initialize: $\mathcal{S} = \{\emptyset\}$, $\omega = 0$.
1: **while** $\omega \leq \omega_c$ **do**
2: For $\mathcal{S}$, compute cut-off estimate $\Omega_k(\mathcal{S})$ and corresponding smoothest signal $\phi_k^* \in L_2(\mathcal{S}^c)$ using Proposition 1.
3: $\omega \leftarrow \Omega_k(\mathcal{S})$, $v \leftarrow \arg\max_i (\phi_k^*(i))^2$.
4: $\mathcal{S} \leftarrow \mathcal{S} \cup v$.
5: **end while**
6: $\mathcal{S}_{\text{est}} \leftarrow \mathcal{S}$.

$$\mathbf{M}_k^\alpha(\mathbf{t}) \triangleq \boldsymbol{\mathcal{L}}^k + \alpha \, \boldsymbol{\mathcal{D}}(\mathbf{t}), \quad k \in \mathbb{Z}^+, \alpha \geq 0, \mathbf{t} \in \mathbb{R}^N \tag{10}$$

where $\boldsymbol{\mathcal{D}}(\mathbf{t})$ is a diagonal matrix with $\mathbf{t}$ on its diagonal. Let $\lambda_k^\alpha(\mathbf{t})$ denote the smallest eigenvalue of $\mathbf{M}_k^\alpha(\mathbf{t})$ and let $\mathbf{1}_\mathcal{S} : \mathcal{V} \to \{0, 1\}$ denote the indicator function for the subset $\mathcal{S}$ (i.e. $\mathbf{1}(\mathcal{S}) = \mathbf{1}$ and $\mathbf{1}(\mathcal{S}^c) = \mathbf{0}$). Then, one has

$$\lambda_k^\alpha(\mathbf{1}_\mathcal{S}) = \inf_{\mathbf{x}} \left(\frac{\mathbf{x}^t \boldsymbol{\mathcal{L}}^k \mathbf{x}}{\mathbf{x}^t \mathbf{x}} + \alpha \frac{\mathbf{x}(\mathcal{S})^t \mathbf{x}(\mathcal{S})}{\mathbf{x}^t \mathbf{x}} \right) \tag{11}$$

Note that the right hand side of the equation above is simply an *unconstrained regularization* of the constrained optimization problem in (5). When $\alpha \gg 1$, the components $\mathbf{x}(\mathcal{S})$ are highly penalized during minimization. Thus, if $\mathbf{x}_k^\alpha(\mathbf{1}_\mathcal{S})$ is the minimizer in (11), then $[\mathbf{x}_k^\alpha(\mathbf{1}_\mathcal{S})](\mathcal{S}) \to \mathbf{0}$, i.e. the values on nodes $\mathcal{S}$ tend to be very small. Therefore, for $\alpha \gg 1$, we have

$$\phi_k^* \approx \mathbf{x}_k^\alpha(\mathbf{1}_\mathcal{S}), \quad \Omega_k(\mathcal{S}) \approx \left(\lambda_k^\alpha(\mathbf{1}_\mathcal{S})\right)^{1/k} \tag{12}$$

By introducing $\lambda_k^\alpha(\mathbf{t})$, we have essentially allowed two relaxations in the computation of $\Omega_k(\mathcal{S})$. Firstly, rather than strictly forcing values of ϕ on $\mathcal{S}$ to be zero, we allow them to be vanishingly small. Secondly, we allow a *binary relaxation* in the argument $\mathbf{t}$, similar to the relaxation used in *graph cut* problems, to understand the variation in $\lambda_k^\alpha(\mathbf{t})$ with $\mathbf{t}$. These relaxations circumvent the combinatorial nature of our problem and have been used earlier to study graph partitioning based on Dirichlet eigenvalues [15, 16]. The effect of adding a node to $\mathcal{S}$ on $\Omega_k(\mathcal{S})$ at each step can now be understood by observing the behavior of the gradient vector $\nabla_\mathbf{t} \lambda_k^\alpha(\mathbf{t})$ at $\mathbf{t} = \mathbf{1}_\mathcal{S}$. Let $\mathbf{x}_k^\alpha(\mathbf{t})$ be the normalized eigenvector of $\mathbf{M}_k^\alpha(\mathbf{t})$ corresponding to $\lambda_k^\alpha(\mathbf{t})$, then

$$\mathbf{M}_k^\alpha(\mathbf{t}) \mathbf{x}_k^\alpha(\mathbf{t}) = \lambda_k^\alpha(\mathbf{t}) \mathbf{x}_k^\alpha(\mathbf{t}) \tag{13}$$

Differentiating both sides with respect to the i^{th} component of $\mathbf{t}$, then multiplying both sides with $\mathbf{x}_k^\alpha(\mathbf{t})^t$, we get

$$\frac{d\lambda_k^\alpha(\mathbf{t})}{d\mathbf{t}(i)} = \mathbf{x}_k^\alpha(\mathbf{t})^t \frac{d\mathbf{M}_k^\alpha(\mathbf{t})}{d\mathbf{t}(i)} \mathbf{x}_k^\alpha(\mathbf{t}) \tag{14}$$

$$= \mathbf{x}_k^\alpha(\mathbf{t})^t \frac{d\left(\alpha \boldsymbol{\mathcal{D}}(\mathbf{t})\right)}{d\mathbf{t}(i)} \mathbf{x}_k^\alpha(\mathbf{t}) = \alpha \left([\mathbf{x}_k^\alpha(\mathbf{t})](i)\right)^2 \tag{15}$$

where $[\mathbf{x}_k^\alpha(\mathbf{t})](i)$ denotes the i^{th} component of the eigenvector $\mathbf{x}_k^\alpha(\mathbf{t})$. Therefore, at point $\mathbf{t}$, $\Delta\lambda_k^\alpha \approx \alpha \left([\mathbf{x}_k^\alpha(\mathbf{t})](i)\right)^2 \Delta\mathbf{t}(i)$. Substituting $\mathbf{t} = \mathbf{1}_\mathcal{S}$ and using equation (12), we have

$$\left.\frac{d\lambda_k^\alpha(\mathbf{t})}{d\mathbf{t}(i)}\right|_{\mathbf{t}=\mathbf{1}_\mathcal{S}} = \alpha \left([\mathbf{x}_k^\alpha(\mathbf{1}_\mathcal{S})](i)\right)^2 \approx \alpha (\phi_k^*(i))^2 \tag{16}$$

The equation above gives us the desired greedy heuristic - starting with an empty $\mathcal{S}$ (i.e. $\mathbf{1}_\mathcal{S} = \mathbf{0}$), if at each step, we include the

node on which the smoothest signal $\phi_k^* \in L_2(\mathcal{S}^c)$ has maximum energy (i.e. $\mathbf{1}_{\mathcal{S}}(i) \leftarrow 1, i = \arg\max_j (\phi_k^*(j))^2$), then $\lambda_k^\alpha(\mathbf{t})$ and in effect, the cut-off estimate $\Omega_k(\mathcal{S})$, tend to increase maximally. We summarize the method for estimating $\mathcal{S}_{\text{opt}}$ in Algorithm 1.

We note that in the algorithm, computing the first eigen-pair of $(\boldsymbol{\mathcal{L}}^k)_{\mathcal{S}^c}$ is the most complex step for each iteration. There are many efficient ways of computing the smallest eigen-pair of a matrix [17], thus allowing the algorithm to be practically feasible. The number of iterations to find $\mathcal{S}^c$ scales linearly with the number of eigenvalues (K_c) below ω_c. Also, with each iteration, the set $\mathcal{S}^c$ shrinks, giving us a bonus reduction of complexity.

Since, $\Omega_k(\mathcal{S})$ underestimates the true cut-off, we may end up with a larger $\mathcal{S}$ than necessary. Another issue is that in the early iterations, using a large k to estimate $\Omega_k(\mathcal{S})$ and ϕ_k^* may result in numerical instabilities because the set $\mathcal{S}$ is small. A solution for this would be to adaptively increase k over the course of iterations. One can also consider numerically stable alternatives to the optimization problem in (5).

3. EXPERIMENTS

In this section, we present simple examples to test the effectiveness of the proposed sampling theory. For our experiments, we first generate four varieties of simple, connected, and undirected graphs:

- G_1: Bipartite graph $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ with $N = 100, |\mathcal{V}_1| = 40, |\mathcal{V}_2| = 60$, random weights $w_{ij} \sim U(0,1)$, sparsified using k-NN, $k = 6$. Nodes in $\mathcal{V}_1$ are numbered 1 to 40 while nodes in $\mathcal{V}_2$ are numbered 41 to 100.
- G_2: Regular cyclic graph with $N = 100$, number of neighbors $K = 8$ and weights $\propto 1/\text{distance}$ (assuming evenly spaced and numbered nodes on a circle).
- G_3: *Erdős-Rényi* random graph with $N = 100, p = 0.2$, weights $w_{ij} \sim U(0,1)$ if i and j are connected.
- G_4: *Watts-Strogatz* 'small-world' model [18], unweighted, with underlying lattice graph $N = 100, K = 8$ and rewiring probability $\beta = 0.1$.

In the first experiment, we randomly generate a subset of nodes $\mathcal{S}$ of certain size for each graph and compare the cut-off estimate from Proposition 1 to that of Theorem 1 [10]. The results are summarized in Table 1. In all cases, it is observed that a higher value of k (> 2) gives better results. The behavior of the cut-off estimate versus k for the particular case of G_1 is shown in Figure 1. Note that $\Omega_k(\mathcal{S})$ increases monotonically with k. In practice, we observe that the computation becomes numerically unstable for large k. For the second experiment, we fix a signal bandwidth $\omega_c = 0.8$ for each graph and estimate $\mathcal{S}_{\text{opt}}$ using Algorithm 1. The number of nodes required for sampling is summarized in Table 2. We note that as k increases, $|\mathcal{S}_{\text{est}}|$ approaches K_c.

For our third experiment, we consider the problem of choosing a subset of nodes $\mathcal{S}$ of given size $|\mathcal{S}| = K$ that maximizes the cut-off frequency. This problem is potentially useful for designing critically sampled filter-banks on graphs. We halt the iterations of our algorithm once a subset of required size is obtained. The indicator function of the subset for two graphs G_1 and G_2 is plotted in Figure 2. For the bipartite graph G_1 with $|\mathcal{S}_{\text{opt}}| = |\mathcal{V}_1| = 40$, we observe from Figure 2(a) that our algorithm successfully chooses most nodes from one of the two colored partitions, in accordance with earlier results [19]. In the cyclic graph G_2, we downsample by a factor of two with $|\mathcal{S}_{\text{opt}}| = N/2 = 50$ and observe that roughly alternate nodes are selected (Figure 2(b)). This is expected because the eigenvector basis for cyclic graphs is the DFT basis and selecting alternate nodes guarantees unique recovery of signals spanned by the first $N/2$ eigenvectors.

Table 1. Cut-off frequency estimates for a given subset of nodes.

Graph	$\lvert\mathcal{S}\rvert$	Theorem 1 ($k=2$)	Proposed method		
			$k=6$	$k=12$	$k=18$
G_1	25	0.2815	0.4070	0.4684	0.5040
G_2	40	0.1236	0.3077	0.4696	0.5427
G_3	25	0.4643	0.6292	0.6756	0.7029
G_4	25	0.2313	0.4716	0.6469	0.7106

Table 2. Estimated sampling set $\mathcal{S}_{\text{est}}$ for bandwidth $\omega_c = 0.8$.

Graph	K_c	$\lvert\mathcal{S}_{\text{est}}\rvert$ using Algorithm 1		
		$k=6$	$k=12$	$k=18$
G_1	36	42	41	41
G_2	41	49	46	45
G_3	24	42	35	32
G_4	19	30	24	22

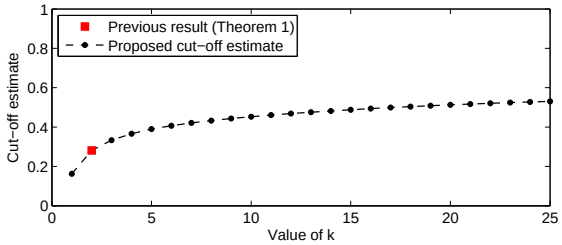

Fig. 1. Behavior of cut-off estimate $\Omega_k(\mathcal{S})$ with k for G_1.

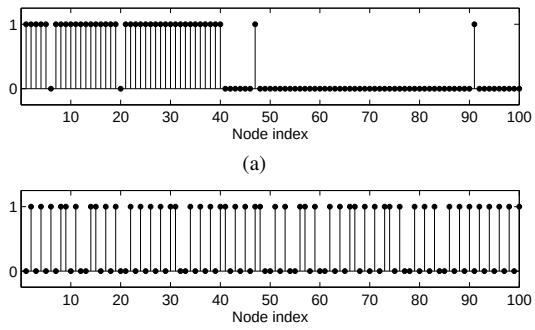

Fig. 2. Indicator functions of $\mathcal{S}_{\text{opt}}$ with $k = 18$ for (a) bipartite graph G_1, $|\mathcal{S}_{\text{opt}}| = |\mathcal{V}_1| = 40$ (b) cyclic graph G_2, $|\mathcal{S}_{\text{opt}}| = N/2 = 50$.

4. CONCLUSIONS

In this paper, we considered the theory of sampling for signals defined on arbitrary graphs using spectral graph theory. Employing the concept of uniqueness sets, we proposed methods to compute the analogs of "Nyquist frequency" and "Nyquist rate" in the context of graph signals. Our results are both novel and practical, and provide a trade-off between accuracy and complexity that one can exploit based on the application at hand. Future work would involve considering robustness to noise in sampling. Another direction would be to look at sampling sets for arbitrary bands of frequencies. This could potentially help us design critically sampled filter-banks on arbitrary graphs.

5. REFERENCES

[1] M. Belkin and P. Niyogi, "Semi-supervised learning on Riemannian manifolds," *Machine Learning*, vol. 56, pp. 209–239, 2004.

[2] X. Zhu and M. Rabbat, "Graph spectral compressed sensing for sensor networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 2865–2868.

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.

[4] F. Zhang and E. R. Hancock, "Graph spectral image smoothing using the heat kernel," *Pattern Recognition*, vol. 41, no. 11, pp. 3328 – 3342, 2008.

[5] S. K. Narang, Y. H. Chao, and A. Ortega, "Graph-wavelet filterbanks for edge-aware image processing," in *Statistical Signal Processing Workshop (SSP), 2012 IEEE*, 2012, pp. 141–144.

[6] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *Signal Processing Magazine, IEEE*, vol. 30, no. 3, pp. 83–98, 2013.

[7] K. Gröchenig, "A discrete theory of irregular sampling," *Linear Algebra and its applications*, vol. 193, pp. 129–150, 1993.

[8] S. K. Narang and A. Ortega, "Downsampling graphs using spectral theory," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, 2011, pp. 4208–4211.

[9] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[10] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 5445–5449.

[11] D. I. Shuman, M. J. Faraji, and P. Vandergheynst, "A framework for multiscale transforms on graphs," Aug. 2013, arXiv:1308.4942 [cs.IT].

[12] F. R. K. Chung, *Spectral graph theory*, vol. 92, CBMS Regional Conference Series in Mathematics, AMS, 1997.

[13] I. Pesenson, "Variational splines and Paley-Wiener spaces on combinatorial graphs," *Constructive Approximation*, vol. 29, no. 1, pp. 1–21, 2009.

[14] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," in *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, Dec 2013, pp. 491–494.

[15] B. Osting, C. D. White, and E. Oudet, "Minimal Dirichlet energy partitions for graphs," Aug. 2013, arXiv:1308.4915 [math.OC].

[16] B. Bourdin, D. Bucur, and E. Oudet, "Optimal partitions for eigenvalues," *SIAM Journal on Scientific Computing*, vol. 31, no. 6, pp. 4100–4114, 2010.

[17] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, SIAM, second edition, 2011.

[18] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[19] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *Signal Processing, IEEE Transactions on*, vol. 60, no. 6, pp. 2786–2799, 2012.